

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Reyes, Gregory R
Yarbough, Patrice O
Bradley, Daniel W
Krawczynski, Krzysztof Z
Tam, Albert
Fry, Kirk E
- (ii) TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
Non-A/Non-B Hepatitis Viral Agent
- (iii) NUMBER OF SEQUENCES: 20
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Dehlinger & Associates
(B) STREET: 350 Cambridge Avenue, Suite 250
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94306
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 09/128,275
(B) FILING DATE: 03-AUG-1998
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/279,823
(B) FILING DATE: 25-JUL-1994
- (viii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/681,078
(B) FILING DATE: 05-APR-1991
- (ix) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/505,888
(B) FILING DATE: 05-APR-1990
- (x) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/420,921
(B) FILING DATE: 13-OCT-1989
- (xi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/367,486
(B) FILING DATE: 16-JUN-1989
- (xii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/336,672
(B) FILING DATE: 11-APR-1989
- (xiii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/208,997
(B) FILING DATE: 17-JUN-1988

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Petithory, Joanne R.
(B) REGISTRATION NUMBER: 42,995
(C) REFERENCE/DOCKET NUMBER: 4600-0183.24

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (650) 324-0880
(B) TELEFAX: (650) 324-0960

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1295 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,
forward sequence

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1293

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..1294

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..1295

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGACCTGTCC CTGTTGCAGC TGTTCTACCA CCCTGCCCCG AGCTCGAACA GGGCCTTCTC	60
TACCTGCCCC AGGAGCTCAC CACCTGTGAT AGTGTCTGTA CATTGAATT AACAGACATT	120
GTGCACTGCC GCATGGCCGC CCCGAGCCAG CGCAAGGCCG TGCTGTCCAC ACTCGTGGGC	180
CGCTACGGCG GTCGCACAAA GCTCTACAAT GCTTCCCACT CTGATGTTTCG CGACTCTCTC	240
GCCCGTTTTA TCCCGGCCAT TGGCCCCGTA CAGGTTACAA CTTGTGAATT GTACGAGCTA	300
GTGGAGGCCA TGGTCGAGAA GGGCCAGGAT GGCTCCGCCG TCCTTGAGCT TGATCTTTGC	360
AACCGTGACG TGTCCAGGAT CACCTTCTTC CAGAAAGATT GTAACAAGTT CACCACAGGT	420
GAGACATTG CCCATGGTAA AGTGGGCCAG GGCATCTCGG CCTGGAGCAA GACCTTCTGC	480
GCCCTCTTTG GCCCTTGTTT CCGCGCTATT GAGAAGGCTA TTCTGGCCCT GTCCTCTCAG	540
GGTGTGTTTT ACGGTGATGC CTTTGATGAC ACCGTCTTCT CGGCGGCTGT GGCCGCAGCA	600

AAGGCATCCA TGGTGTTTGA GAATGACTTT TCTGAGTTTG ACTCCACCCA GAATAACTTT 660
 TCTCTGGGTC TAGAGTGTGC TATTATGGAG GAGTGTGGGA TGCCGCAGTG GCTCATCCGC 720
 CTGTATCACC TTATAAGGTC TGGGTGGATC TTGCAGGCCC CGAAGGAGTC TCTGCGAGGG 780
 TTTTGGAAGA AACACTCCGG TGAGCCCGGC ACTCTTCTAT GGAATACTGT CTGGAATATG 840
 GCCGTTATTA CCCACTGTTA TGACTTCCGC GATTTTCAGG TGGCTGCCTT TAAAGGTGAT 900
 GATTCGATAG TGCTTTGCAG TGAGTATCGT CAGAGTCCAG GAGCTGCTGT CCTGATCGCC 960
 GGCTGTGGCT TGAAGTTGAA GGTAGATTTT CGCCCGATCG GTTTGTATGC AGGTGTTGTG 1020
 GTGGCCCCCG GCCTTGGCGC GCTCCCTGAT GTTGTGCGCT TCGCCGGCCG GCTTACCGAG 1080
 AAGAATTGGG GCCCTGGCCC TGAGCGGGCG GAGCAGCTCC GCCTCGCTGT TAGTGATTTC 1140
 CTCCGCAAGC TCACGAATGT AGCTCAGATG TGTGTGGATG TTGTTTCCCG TGTTTATGGG 1200
 GTTTCCCCTG GACTCGTTCA TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG 1260
 GCACATTTC A CTGAGTCAGT AAAACCAGTG CTCGA 1295

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 431 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Arg	Pro	Val	Pro	Val	Ala	Ala	Val	Leu	Pro	Pro	Cys	Pro	Glu	Leu	Glu	1	5	10	15
Gln	Gly	Leu	Leu	Tyr	Leu	Pro	Gln	Glu	Leu	Thr	Thr	Cys	Asp	Ser	Val	20	25	30	
Val	Thr	Phe	Glu	Leu	Thr	Asp	Ile	Val	His	Cys	Arg	Met	Ala	Ala	Pro	35	40	45	
Ser	Gln	Arg	Lys	Ala	Val	Leu	Ser	Thr	Leu	Val	Gly	Arg	Tyr	Gly	Gly	50	55	60	
Arg	Thr	Lys	Leu	Tyr	Asn	Ala	Ser	His	Ser	Asp	Val	Arg	Asp	Ser	Leu	65	70	75	80
Ala	Arg	Phe	Ile	Pro	Ala	Ile	Gly	Pro	Val	Gln	Val	Thr	Thr	Cys	Glu	85	90	95	
Leu	Tyr	Glu	Leu	Val	Glu	Ala	Met	Val	Glu	Lys	Gly	Gln	Asp	Gly	Ser	100	105	110	
Ala	Val	Leu	Glu	Leu	Asp	Leu	Cys	Asn	Arg	Asp	Val	Ser	Arg	Ile	Thr	115	120	125	
Phe	Phe	Gln	Lys	Asp	Cys	Asn	Lys	Phe	Thr	Thr	Gly	Glu	Thr	Ile	Ala	130	135	140	

09851410-050701

His	Gly	Lys	Val	Gly	Gln	Gly	Ile	Ser	Ala	Trp	Ser	Lys	Thr	Phe	Cys	145	150	155	160
Ala	Leu	Phe	Gly	Pro	Trp	Phe	Arg	Ala	Ile	Glu	Lys	Ala	Ile	Leu	Ala	165	170		175
Leu	Leu	Pro	Gln	Gly	Val	Phe	Tyr	Gly	Asp	Ala	Phe	Asp	Asp	Thr	Val	180	185		190
Phe	Ser	Ala	Ala	Val	Ala	Ala	Ala	Lys	Ala	Ser	Met	Val	Phe	Glu	Asn	195	200		205
Asp	Phe	Ser	Glu	Phe	Asp	Ser	Thr	Gln	Asn	Asn	Phe	Ser	Leu	Gly	Leu	210	215		220
Glu	Cys	Ala	Ile	Met	Glu	Glu	Cys	Gly	Met	Pro	Gln	Trp	Leu	Ile	Arg	225	230		235
Leu	Tyr	His	Leu	Ile	Arg	Ser	Ala	Trp	Ile	Leu	Gln	Ala	Pro	Lys	Glu	245	250		255
Ser	Leu	Arg	Gly	Phe	Trp	Lys	Lys	His	Ser	Gly	Glu	Pro	Gly	Thr	Leu	260	265		270
Leu	Trp	Asn	Thr	Val	Trp	Asn	Met	Ala	Val	Ile	Thr	His	Cys	Tyr	Asp	275	280		285
Phe	Arg	Asp	Phe	Gln	Val	Ala	Ala	Phe	Lys	Gly	Asp	Asp	Ser	Ile	Val	290	295		300
Leu	Cys	Ser	Glu	Tyr	Arg	Gln	Ser	Pro	Gly	Ala	Ala	Val	Leu	Ile	Ala	305	310		315
Gly	Cys	Gly	Leu	Lys	Leu	Lys	Val	Asp	Phe	Arg	Pro	Ile	Gly	Leu	Tyr	325	330		335
Ala	Gly	Val	Val	Val	Ala	Pro	Gly	Leu	Gly	Ala	Leu	Pro	Asp	Val	Val	340	345		350
Arg	Phe	Ala	Gly	Arg	Leu	Thr	Glu	Lys	Asn	Trp	Gly	Pro	Gly	Pro	Glu	355	360		365
Arg	Ala	Glu	Gln	Leu	Arg	Leu	Ala	Val	Ser	Asp	Phe	Leu	Arg	Lys	Leu	370	375		380
Thr	Asn	Val	Ala	Gln	Met	Cys	Val	Asp	Val	Val	Ser	Arg	Val	Tyr	Gly	385	390		395
Val	Ser	Pro	Gly	Leu	Val	His	Asn	Leu	Ile	Gly	Met	Leu	Gln	Ala	Val	405	410		415
Ala	Asp	Gly	Lys	Ala	His	Phe	Thr	Glu	Ser	Val	Lys	Pro	Val	Leu		420	425		430

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: linker - top (5') sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGAATTCGCG GCCGCTCG

18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: linker - bottom (3') sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGAGCGGCCG CGAATTCCTT

20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1295 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,
 reverse sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCGAGCACTG GTTTTACTGA CTCAGTGAAA TGTGCCTTGC CATCAGCAAC AGCCTGTAGC
 ATGCCAATCA GGTTATGAAC GAGTCCAGGG GAAACCCCAT AAACACGGGA AACAAACATCC

60

120

ACACACATCT GAGCTACATT CGTGAGCTTG CGGAGGAAAT CACTAACAGC GAGGCGGAGC	180
TGCTCCGCCC GCTCAGGGCC AGGGCCCCAA TTCTTCTCGG TAAGCCGGCC GGCGAAGCGC	240
ACAACATCAG GGAGCGCGCC AAGGCCGGGG GCCACCACAA CACCTGCATA CAAACCGATC	300
GGGCGGAAAT CTACCTTCAA CTTCAAGCCA CAGCCGGCGA TCAGGACAGC AGCTCCTGGA	360
CTCTGACGAT ACTCACTGCA AAGCACTATC GAATCATCAC CTTTAAAGGC AGCCACCTGA	420
AAATCGCGGA AGTCATAACA GTGGGTAATA ACGGCCATAT TCCAGACAGT ATTCCATAGA	480
AGAGTGCCGG GCTCACCGBA GTGTTTCTTC CAAAACCCTC GCAGAGACTC CTTCGGGGCC	540
TGCAAGATCC ACGCAGACCT TATAAGGTGA TACAGGCGGA TGAGCCACTG CGGCATCCCA	600
CACTCCTCCA TAATAGCACA CTCTAGACCC AGAGAAAAGT TATTCTGGGT GGAGTCAAAC	660
TCAGAAAAGT CATTCTCAA CACCATGGAT GCCTTTGCTG CGGCCACAGC CGCCGAGAAG	720
ACGGTGTCTA CAAAGGCATC ACCGTAAAAC ACACCCTGAG GGAGCAGGGC CAGAATAGCC	780
TTCTCAATAG CGCGGAACCA AGGGCCAAAG AGGGCGCAGA AGGTCTTGCT CCAGGCCGAG	840
ATGCCCTGGC CCACTTTACC ATGGGCAATG GTCTCACCTG TGGTGAACCT GTTACAATCT	900
TTCTGGAAGA AGGTGATCCT GGACACGTCA CGGTTGCAAA GATCAAGCTC AAGGACGGCG	960
GAGCCATCCT GGCCCTTCTC GACCATGGCC TCCACTAGCT CGTACAATTC ACAAGTTGTA	1020
ACCTGTACGG GGCCAATGGC CGGGATAAAA CGGGCGAGAG AGTCGCGAAC ATCAGAGTGG	1080
GAAGCATTGT AGAGCTTTGT GCGACCGCCG TAGCGGCCCA CGAGTGTGGA CAGCACGGCC	1140
TTGCGCTGGC TCGGGGCGGC CATGCGGCAG TGCACAATGT CTGTTAATTC AAATGTTACG	1200
AACTATCAC AGGTGGTGAG CTCCTGGGGC AGGTAGAGAA GGCCCTGTTC GAGCTCGGGG	1260
CAGGGTGGTA GAACAGCTGC AACAGGGACA GGTCT	1295

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7195 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (C) INDIVIDUAL ISOLATE: HEV - Burma strain
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 28..5106

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 5147..7126

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 5106..5474

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGCAGACCA CATATGTGGT CGATGCCATG GAGGCCCATC AGTTTATTAA GGCTCCTGGC	60
ATCACTACTG CTATTGAGCA GGCTGCTCTA GCAGCGGCCA ACTCTGCCCT GGCGAATGCT	120
GTGGTAGTTA GGCCTTTTCT CTCTCACCAG CAGATTGAGA TCCTCATTAA CCTAATGCAA	180
CCTCGCCAGC TTGTTTTCCG CCCCAGAGTT TTCTGGAATC ATCCCATCCA GCGTGTCATC	240
CATAACGAGC TGGAGCTTTA CTGCCGCGCC CGCTCCGGCC GCTGTCTTGA AATTGGCGCC	300
CATCCCCGCT CAATAAATGA TAATCCTAAT GTGGTCCACC GCTGCTTCCT CCGCCCTGTT	360
GGGCGTGATG TTCAGCGCTG GTATACTGCT CCCACTCGCG GGCCGGCTGC TAATTGCCGG	420
CGTTCGCGCG TCGCGGGCT TCCCGCTGCT GACCGCACTT ACTGCCTCGA CGGGTTTTCT	480
GGCTGTAAC TTTCCGCCGA GACTGGCATC GCCCTCTACT CCCTTCATGA TATGTCACCA	540
TCTGATGTCG CCGAGGCCAT GTTCCGCCAT GGTATGACGC GGCTCTATGC CGCCCTCCAT	600
CTTCCGCCCTG AGGTCCTGCT GCCCCCTGGC ACATATCGCA CCGCATCGTA TTTGCTAATT	660
CATGACGGTA GGC CGTGTGT GGTGACGTAT GAGGGTGATA CTAGTGCTGG TTACAACCAC	720
GATGTCTCCA ACTTGCCTC CTGGATTAGA ACCACCAAGG TTACCGGAGA CCATCCCCCTC	780
GTTATCGAGC GGGTTAGGGC CATTGGCTGC CACTTTGTTC TCTTGCTCAC GGCAGCCCCG	840
GAGCCATCAC CTATGCCTTA TGTTCTTAC CCCCAGTCTA CCGAGGTCTA TGTCGATCG	900
ATCTTCGGCC CGGGTGGCAC CCCTTCCTTA TTCCCAACCT CATGCTCCAC TAAGTCGACC	960
TTCCATGCTG TCCCTGCCCA TATTTGGGAC CGTCTTATGC TGTTGCGGGC CACCTTGGAT	1020
GACCAAGCCT TTTGCTGCTC CCGTTTAATG ACCTACCTTC GCGGCATTAG CTACAAGGTC	1080
ACTGTTGGTA CCCTTGTTGG TAATGAAGGC TGAATGCCT CTGAGGACGC CCTCACAGCT	1140
GTTATCACTG CCGCTACCT TACCATTTGC CACCAGCGGT ATCTCCGCAC CCAGGCTATA	1200
TCCAAGGGGA TCGTCGTCT GGAACGGGAG CATGCCGAGA AGTTTATAAC ACGCCTCTAC	1260
AGCTGGCTCT TCGAGAAGTC CGGCCGTGAT TACATCCCTG GCCGTCAGTT GGAGTTCTAC	1320
GCCCAAGTGA GGCCTGGCT CTCCGCCGGC TTTCATCTTG ATCCACGGGT GTTGGTTTTT	1380
GACGAGTCGG CCCCCTGCCA TTGTAGGACC GCGATCCGTA AGGCGCTCTC AAAGTTTTGC	1440
TGCTTCATGA AGTGGCTTGG TCAGGAGTGC ACCTGCTTCC TTCAGCCTGC AGAAGGCGCC	1500
GTCGGCGACC AGGGTCATGA TAATGAAGCC TATGAGGGGT CCGATGTTGA CCCTGCTGAG	1560

TCCGCCATTA GTGACATATC TGGGTCCTAT GTCGTCCCTG GCACTGCCCT CCAACCGCTC	1620
TACCAGGCCC TCGATCTCCC CGCTGAGATT GTGGCTCGCG CGGGCCGGCT GACCGCCACA	1680
GTAAAGGTCT CCCAGGTCGA TGGGCGGATC GATTGCGAGA CCCTTCTTGG TAACAAAACC	1740
TTTCGCACGT CGTTCGTTGA CGGGGCGGTC TTAGAGACCA ATGGCCCAGA GCGCCACAAT	1800
CTCTCCTTCG ATGCCAGTCA GAGCACTATG GCCGCTGGCC CTTTCAGTCT CACCTATGCC	1860
GCCTCTGCAG CTGGGCTGGA GGTGCGCTAT GTTGCTGCCG GGCTTGACCA TCGGGCGGTT	1920
TTTGCCCCCG GTGTTTCACC CCGGTCAGCC CCCGGCGAGG TTACCGCCTT CTGCTCTGCC	1980
CTATACAGGT TTAACCGTGA GGCCAGCGC CATTCGCTGA TCGGTAACCT ATGGTTCCAT	2040
CCTGAGGGAC TCATTGGCCT CTTGCCCCG TTTTCGCCCC GGCATGTTTG GGAGTCGGCT	2100
AATCCATTCT GTGGCGAGAG CACACTTTAC ACCCGTACTT GGTCGGAGGT TGATGCCGTC	2160
TCTAGTCCAG CCCGGCCTGA CTTAGGTTTT ATGTCTGAGC CTTCTATACC TAGTAGGGCC	2220
GCCACGCCTA CCCTGGCGGC CCCTCTACCC CCCCCTGCAC CGGACCCCTC CCCCCCTCCC	2280
TCTGCCCCCG CGCTTGCTGA GCCGGCTTCT GCGGCTACCG CCGGGGCCCC GGCCATAACT	2340
CACCAGACGG CCCGGCACCG CCGCCTGCTC TTCACCTACC CGGATGGCTC TAAGGTATTC	2400
GCCGGCTCGC TGTTTCGAGTC GACATGCACG TGGCTCGTTA ACGCGTCTAA TGTTGACCAC	2460
CGCCCTGGCG GCGGGCTTTG CCATGCATTT TACCAAAGGT ACCCGCCTC CTTTGATGCT	2520
GCCTCTTTTG TGATGCGCGA CGGCGCGGCC GCGTACACAC TAACCCCCCG GCCAATAATT	2580
CACGCTGTCTG CCCCTGATTA TAGGTTGGAA CATAACCCAA AGAGGCTTGA GGCTGCTTAT	2640
CGGGAAACTT GCTCCCGCCT CGGCACCGCT GCATACCCGC TCCTCGGGAC CGGCATATAC	2700
CAGGTGCCGA TCGGCCCCAG TTTGACGCC TGGGAGCGGA ACCACCGCCC CGGGGATGAG	2760
TTGTACCTTC CTGAGCTTGC TGCCAGATGG TTTGAGGCCA ATAGGCCGAC CCGCCCGACT	2820
CTCACTATAA CTGAGGATGT TGCACGGACA GCGAATCTGG CCATCGAGCT TGA CTGAGCC	2880
ACAGATGTCTG GCCGGGCCTG TGCCGGCTGT CGGGTCACCC CCGGCGTTGT TCAGTACCAG	2940
TTTACTGCAG GTGTGCCTGG ATCCGGCAAG TCCCCTCTA TCACCCAAGC CGATGTGGAC	3000
GTTGTCGTGG TCCCGACGCG TGAGTTGCGT AATGCCTGGC GCCGTCGCGG CTTTGCTGCT	3060
TTTACCCCGC ATACTGCCGC CAGAGTCACC CAGGGGCGCC GGGTTGTCAT TGATGAGGCT	3120
CCATCCCTCC CCCCTCACCT GCTGCTGCTC CACATGCAGC GGGCCGCCAC CGTCCACCTT	3180
CTTGCGGACC CGAACCAGAT CCCAGCCATC GACTTTGAGC ACGCTGGGCT CGTCCCCGCC	3240
ATCAGGCCCC ACTTAGGCCC CACCTCCTGG TGGCATGTTA CCCATCGCTG GCCTGCGGAT	3300
GTATGCGAGC TCATCCGTGG TGCATACCCC ATGATCCAGA CCACTAGCCG GGTTCCTCGT	3360
TCGTTGTTCT GGGGTGAGCC TGCCGTCGGG CAGAACTAG TGTTCACCCA GGCGGCCAAG	3420

CCCGCCAACC	CCGGCTCAGT	GACGGTCCAC	GAGGCGCAGG	GCGCTACCTA	CACGGAGACC	3480
ACTATTATTG	CCACAGCAGA	TGCCCCGGGGC	CTTATTCAGT	CGTCTCGGGC	TCATGCCATT	3540
GTTGCTCTGA	CGCGCCACAC	TGAGAAGTGC	GTCATCATTG	ACGCACCAGG	CCTGCTTCGC	3600
GAGGTGGGCA	TCTCCGATGC	AATCGTTAAT	AACTTTTTCC	TCGCTGGTGG	CGAAATTGGT	3660
CACCAGCGCC	CATCAGTTAT	TCCCCGTGGC	AACCCTGACG	CCAATGTTGA	CACCCTGGCT	3720
GCCTTCCCGC	CGTCTTGCCA	GATTAGTGCC	TTCCATCAGT	TGGCTGAGGA	GCTTGGCCAC	3780
AGACCTGTCC	CTGTTGCAGC	TGTTCTACCA	CCCTGCCCCG	AGCTCGAACA	GGGCCTTCTC	3840
TACCTGCCCC	AGGAGCTCAC	CACCTGTGAT	AGTGTGCTAA	CATTTGAATT	AACAGACATT	3900
GTGCACTGCC	GCATGGCCGC	CCCAGCCAG	CGCAAGGCCG	TGCTGTCCAC	ACTCGTGGGC	3960
CGCTACGGCG	GTCGCACAAA	GCTCTACAAT	GCTTCCCACT	CTGATGTTTC	CGACTCTCTC	4020
GCCCCGTTTTA	TCCCCGCCAT	TGGCCCCGTA	CAGGTTACAA	CTTGTGAATT	GTACGAGCTA	4080
GTGGAGGCCA	TGGTCGAGAA	GGGCCAGGAT	GGCTCCGCCG	TCCTTGAGCT	TGATCTTTGC	4140
AACCGTGACG	TGTCCAGGAT	CACCTTCTTC	CAGAAAGATT	GTAACAAGTT	CACCACAGGT	4200
GAGACCATTG	CCCATGGTAA	AGTGGGCCAG	GGCATCTCGG	CCTGGAGCAA	GACCTTCTGC	4260
GCCCTCTTTG	GCCCTTGGTT	CCGCGCTATT	GAGAAGGCTA	TTCTGGCCCT	GCTCCCTCAG	4320
GGTGTGTTTT	ACGGTGATGC	CTTTGATGAC	ACCGTCTTCT	CGGCGGCTGT	GGCCGAGCA	4380
AAGGCATCCA	TGGTGTTTGA	GAATGACTTT	TCTGAGTTTG	ACTCCACCCA	GAATAACTTT	4440
TCTCTGGGTC	TAGAGTGTGC	TATTATGGAG	GAGTGTGGGA	TGCCGCAGTG	GCTCATCCGC	4500
CTGTATCACC	TTATAAGGTC	TGCGTGGATC	TTGCAGGCCC	CGAAGGAGTC	TCTGCGAGGG	4560
TTTTGGAAGA	AACACTCCGG	TGAGCCCGGC	ACTCTTCTAT	GGAATACTGT	CTGGAATATG	4620
GCCGTTATTA	CCCACTGTTA	TGACTTCCGC	GATTTTCAGG	TGGCTGCCTT	TAAAGGTGAT	4680
GATTCGATAG	TGCTTTGCAG	TGAGTATCGT	CAGAGTCCAG	GAGCTGCTGT	CCTGATCGCC	4740
GGCTGTGGCT	TGAAGTTGAA	GGTAGATTTT	CGCCCGATCG	GTTTGTATGC	AGGTGTTGTG	4800
GTGGCCCCCG	GCCTTGCGCG	GCTCCCTGAT	GTTGTGCGCT	TCGCCGGCCG	GCTTACCGAG	4860
AAGAATTGGG	GCCCTGGCCC	TGAGCGGGCG	GAGCAGCTCC	GCCTCGCTGT	TAGTGATTTT	4920
CTCCGCAAGC	TCACGAATGT	AGCTCAGATG	TGTGTGGATG	TTGTTTCCCG	TGTTTATGGG	4980
GTTTCCCCTG	GACTCGTTCA	TAACCTGATT	GGCATGCTAC	AGGCTGTTGC	TGATGGCAAG	5040
GCACATTTCA	CTGAGTCAGT	AAAACCAAGT	CTCGACTTGA	CAAATTCAAT	CTTGTGTCGG	5100
GTGGAATGAA	TAACATGTCT	TTTGCTGCGC	CCATGGGTTC	GCGACCATGC	GCCCTCGGCC	5160
TATTTTGTTG	CTGCTCCTCA	TGTTTTTGCC	TATGCTGCCC	GCGCCACCGC	CCGGTCAGCC	5220
GTCTGGCCGC	CGTCGTGGGC	GGCGCAGCGG	CGGTTCCGGC	GGTGGTTTCT	GGGGTGACCG	5280

GGTTGATTCT	CAGCCCTTCG	CAATCCCCTA	TATTCATCCA	ACCAACCCCT	TCGCCCCCGA	5340
TGTCACCGCT	GCGGCCGGGG	CTGGACCTCG	TGTTCCGCAA	CCCGCCCGAC	CACTCGGCTC	5400
CGCTTGGCGT	GACCAGGCCC	AGCGCCCCGC	CGTTGCCTCA	CGTCGTAGAC	CTACCACAGC	5460
TGGGGCCGCG	CCGCTAACCG	CGGTCGCTCC	GGCCCATGAC	ACCCCGCCAG	TGCCTGATGT	5520
CGACTCCCGC	GGCGCCATCT	TGCGCCGGCA	GTATAACCTA	TCAACATCTC	CCCTTACCTC	5580
TTCCGTGGCC	ACCGGCACTA	ACCTGGTTCT	TTATGCCGCC	CCTCTTAGTC	CGCTTTTACC	5640
CCTTCAGGAC	GGCACCAATA	CCCATATAAT	GGCCACGGAA	GCTTCTAATT	ATGCCCAGTA	5700
CCGGGTTGCC	CGTGCCACAA	TCCGTTACCG	CCCCTGGTTC	CCCAATGCTG	TCGGCGGTTA	5760
CGCCATCTCC	ATCTCATTCT	GGCCACAGAC	CACCACCACC	CCGACGTCCG	TTGATATGAA	5820
TTCAATAACC	TCGACGGATG	TTCGTATTTT	AGTCCAGCCC	GGCATAGCCT	CTGAGCTTGT	5880
GATCCCAAGT	GAGCGCCTAC	ACTATCGTAA	CCAAGGCTGG	CGCTCCGTCG	AGACCTCTGG	5940
GGTGGCTGAG	GAGGAGGCTA	CCTCTGGTCT	TGTTATGCTT	TGCATACATG	GCTCACTCGT	6000
AAATTCCCTAT	ACTAATACAC	CCTATACCGG	TGCCCTCGGG	CTGTTGGACT	TTGCCCTTGA	6060
GCTTGAGTTT	CGCAACCTTA	CCCCCGGTAA	CACCAATACG	CGGGTCTCCC	GTTATTCCAG	6120
CACTGCTCGC	CACCGCCTTC	GTCGCGGTGC	GGACGGGACT	GCCGAGCTCA	CCACCACGGC	6180
TGCTACCCGC	TTTATGAAGG	ACCTCTATTT	TACTAGTACT	AATGGTGTCT	GTGAGATCGG	6240
CCGCGGGATA	GCCCTCACCC	TGTTCAACCT	TGCTGACACT	CTGCTTGGCG	GCCTGCCGAC	6300
AGAATTGATT	TCGTGCGCTG	GTGGCCAGCT	GTTCTACTCC	CGTCCCGTTG	TCTCAGCCAA	6360
TGGCGAGCCG	ACTGTTAAGT	TGTATACATC	TGTAGAGAAT	GCTCAGCAGG	ATAAGGGTAT	6420
TGCAATCCCG	CATGACATTG	ACCTCGGAGA	ATCTCGTGTG	GTTATTCAGG	ATTATGATAA	6480
CCAACATGAA	CAAGATCGGC	CGACGCCTTC	TCCAGCCCCA	TCGCGCCCTT	TCTCTGTCCT	6540
TCGAGCTAAT	GATGTGCTTT	GGCTCTCTCT	CACCGCTGCC	GAGTATGACC	AGTCCACTTA	6600
TGGCTCTTCG	ACTGGCCCAG	TTTATGTTTC	TGACTCTGTG	ACCTTGGTTA	ATGTTGCGAC	6660
CGGCGCGCAG	GCCGTTGCCC	GGTCGCTCGA	TTGGACCAAG	GTCACACTTG	ACGGTCGCCC	6720
CCTCTCCACC	ATCCAGCAGT	ACTCGAAGAC	CTTCTTTGTC	CTGCCGCTCC	GCGGTAAGCT	6780
CTCTTTCTGG	GAGGCAGGCA	CAACTAAAGC	CGGGTACCCT	TATAATTATA	ACACCACTGC	6840
TAGCGACCAA	CTGCTTGTCG	AGAATGCCCG	CGGGCACCGG	GTCGCTATTT	CCACTTACAC	6900
CACTAGCCTG	GGTGCTGGTC	CCGTCTCCAT	TTCTGCGGTT	GCCGTTTTAG	CCCCCACTC	6960
TGCGCTAGCA	TTGCTTGAGG	ATACCTTGGA	CTACCCTGCC	CGCGCCCAT	CTTTTGATGA	7020
TTTCTGCCCA	GAGTGCCGCC	CCCTTGGCCT	TCAGGGCTGC	GCTTTCCAGT	CTACTGTGCG	7080
TGAGCTTCAG	CGCCTTAAGA	TGAAGGTGGG	TAAAACTCGG	GAGTTGTAGT	TTATTTGCTT	7140

GTGCCCCCT TCTTCTGTT GCTTATTTCT CATTTCTGCG TTCCGCGCTC CCTGA

7195

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1693 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Glu	Ala	His	Gln	Phe	Ile	Lys	Ala	Pro	Gly	Ile	Thr	Thr	Ala	Ile
1				5					10					15	
Glu	Gln	Ala	Ala	Leu	Ala	Ala	Ala	Asn	Ser	Ala	Leu	Ala	Asn	Ala	Val
			20					25					30		
Val	Val	Arg	Pro	Phe	Leu	Ser	His	Gln	Gln	Ile	Glu	Ile	Leu	Ile	Asn
		35					40					45			
Leu	Met	Gln	Pro	Arg	Gln	Leu	Val	Phe	Arg	Pro	Glu	Val	Phe	Trp	Asn
	50					55					60				
His	Pro	Ile	Gln	Arg	Val	Ile	His	Asn	Glu	Leu	Glu	Leu	Tyr	Cys	Arg
	65				70					75					80
Ala	Arg	Ser	Gly	Arg	Cys	Leu	Glu	Ile	Gly	Ala	His	Pro	Arg	Ser	Ile
				85					90					95	
Asn	Asp	Asn	Pro	Asn	Val	Val	His	Arg	Cys	Phe	Leu	Arg	Pro	Val	Gly
			100					105					110		
Arg	Asp	Val	Gln	Arg	Trp	Tyr	Thr	Ala	Pro	Thr	Arg	Gly	Pro	Ala	Ala
		115					120					125			
Asn	Cys	Arg	Arg	Ser	Ala	Leu	Arg	Gly	Leu	Pro	Ala	Ala	Asp	Arg	Thr
	130					135					140				
Tyr	Cys	Leu	Asp	Gly	Phe	Ser	Gly	Cys	Asn	Phe	Pro	Ala	Glu	Thr	Gly
	145				150					155					160
Ile	Ala	Leu	Tyr	Ser	Leu	His	Asp	Met	Ser	Pro	Ser	Asp	Val	Ala	Glu
				165					170					175	
Ala	Met	Phe	Arg	His	Gly	Met	Thr	Arg	Leu	Tyr	Ala	Ala	Leu	His	Leu
			180					185					190		
Pro	Pro	Glu	Val	Leu	Leu	Pro	Pro	Gly	Thr	Tyr	Arg	Thr	Ala	Ser	Tyr
		195				200						205			
Leu	Leu	Ile	His	Asp	Gly	Arg	Arg	Val	Val	Val	Thr	Tyr	Glu	Gly	Asp
	210					215					220				
Thr	Ser	Ala	Gly	Tyr	Asn	His	Asp	Val	Ser	Asn	Leu	Arg	Ser	Trp	Ile
	225				230					235					240
Arg	Thr	Thr	Lys	Val	Thr	Gly	Asp	His	Pro	Leu	Val	Ile	Glu	Arg	Val

				245						250					255
Arg	Ala	Ile	Gly	Cys	His	Phe	Val	Leu	Leu	Leu	Thr	Ala	Ala	Pro	Glu
			260					265					270		
Pro	Ser	Pro	Met	Pro	Tyr	Val	Pro	Tyr	Pro	Arg	Ser	Thr	Glu	Val	Tyr
		275					280					285			
Val	Arg	Ser	Ile	Phe	Gly	Pro	Gly	Gly	Thr	Pro	Ser	Leu	Phe	Pro	Thr
	290					295					300				
Ser	Cys	Ser	Thr	Lys	Ser	Thr	Phe	His	Ala	Val	Pro	Ala	His	Ile	Trp
305					310					315					320
Asp	Arg	Leu	Met	Leu	Phe	Gly	Ala	Thr	Leu	Asp	Asp	Gln	Ala	Phe	Cys
				325					330					335	
Cys	Ser	Arg	Leu	Met	Thr	Tyr	Leu	Arg	Gly	Ile	Ser	Tyr	Lys	Val	Thr
			340					345					350		
Val	Gly	Thr	Leu	Val	Ala	Asn	Glu	Gly	Trp	Asn	Ala	Ser	Glu	Asp	Ala
		355					360					365			
Leu	Thr	Ala	Val	Ile	Thr	Ala	Ala	Tyr	Leu	Thr	Ile	Cys	His	Gln	Arg
	370					375					380				
Tyr	Leu	Arg	Thr	Gln	Ala	Ile	Ser	Lys	Gly	Met	Arg	Arg	Leu	Glu	Arg
385					390					395					400
Glu	His	Ala	Gln	Lys	Phe	Ile	Thr	Arg	Leu	Tyr	Ser	Trp	Leu	Phe	Glu
				405					410					415	
Lys	Ser	Gly	Arg	Asp	Tyr	Ile	Pro	Gly	Arg	Gln	Leu	Glu	Phe	Tyr	Ala
			420					425					430		
Gln	Cys	Arg	Arg	Trp	Leu	Ser	Ala	Gly	Phe	His	Leu	Asp	Pro	Arg	Val
		435					440					445			
Leu	Val	Phe	Asp	Glu	Ser	Ala	Pro	Cys	His	Cys	Arg	Thr	Ala	Ile	Arg
	450					455					460				
Lys	Ala	Leu	Ser	Lys	Phe	Cys	Cys	Phe	Met	Lys	Trp	Leu	Gly	Gln	Glu
465					470					475					480
Cys	Thr	Cys	Phe	Leu	Gln	Pro	Ala	Glu	Gly	Ala	Val	Gly	Asp	Gln	Gly
				485					490					495	
His	Asp	Asn	Glu	Ala	Tyr	Glu	Gly	Ser	Asp	Val	Asp	Pro	Ala	Glu	Ser
			500					505					510		
Ala	Ile	Ser	Asp	Ile	Ser	Gly	Ser	Tyr	Val	Val	Pro	Gly	Thr	Ala	Leu
		515					520					525			
Gln	Pro	Leu	Tyr	Gln	Ala	Leu	Asp	Leu	Pro	Ala	Glu	Ile	Val	Ala	Arg
	530					535					540				
Ala	Gly	Arg	Leu	Thr	Ala	Thr	Val	Lys	Val	Ser	Gln	Val	Asp	Gly	Arg
545					550					555					560
Ile	Asp	Cys	Glu	Thr	Leu	Leu	Gly	Asn	Lys	Thr	Phe	Arg	Thr	Ser	Phe
				565					570					575	

Val Asp Gly Ala Val Leu Glu Thr Asn Gly Pro Glu Arg His Asn Leu
580 585 590

Ser Phe Asp Ala Ser Gln Ser Thr Met Ala Ala Gly Pro Phe Ser Leu
595 600 605

Thr Tyr Ala Ala Ser Ala Ala Gly Leu Glu Val Arg Tyr Val Ala Ala
610 615 620

Gly Leu Asp His Arg Ala Val Phe Ala Pro Gly Val Ser Pro Arg Ser
625 630 635 640

Ala Pro Gly Glu Val Thr Ala Phe Cys Ser Ala Leu Tyr Arg Phe Asn
645 650 655

Arg Glu Ala Gln Arg His Ser Leu Ile Gly Asn Leu Trp Phe His Pro
660 665 670

Glu Gly Leu Ile Gly Leu Phe Ala Pro Phe Ser Pro Gly His Val Trp
675 680 685

Glu Ser Ala Asn Pro Phe Cys Gly Glu Ser Thr Leu Tyr Thr Arg Thr
690 695 700

Trp Ser Glu Val Asp Ala Val Ser Ser Pro Ala Arg Pro Asp Leu Gly
705 710 715 720

Phe Met Ser Glu Pro Ser Ile Pro Ser Arg Ala Ala Thr Pro Thr Leu
725 730 735

Ala Ala Pro Leu Pro Pro Pro Ala Pro Asp Pro Ser Pro Pro Pro Ser
740 745 750

Ala Pro Ala Leu Ala Glu Pro Ala Ser Gly Ala Thr Ala Gly Ala Pro
755 760 765

Ala Ile Thr His Gln Thr Ala Arg His Arg Arg Leu Leu Phe Thr Tyr
770 775 780

Pro Asp Gly Ser Lys Val Phe Ala Gly Ser Leu Phe Glu Ser Thr Cys
785 790 795 800

Thr Trp Leu Val Asn Ala Ser Asn Val Asp His Arg Pro Gly Gly Gly
805 810 815

Leu Cys His Ala Phe Tyr Gln Arg Tyr Pro Ala Ser Phe Asp Ala Ala
820 825 830

Ser Phe Val Met Arg Asp Gly Ala Ala Ala Tyr Thr Leu Thr Pro Arg
835 840 845

Pro Ile Ile His Ala Val Ala Pro Asp Tyr Arg Leu Glu His Asn Pro
850 855 860

Lys Arg Leu Glu Ala Ala Tyr Arg Glu Thr Cys Ser Arg Leu Gly Thr
865 870 875 880

Ala Ala Tyr Pro Leu Leu Gly Thr Gly Ile Tyr Gln Val Pro Ile Gly
885 890 895

Pro Ser Phe Asp Ala Trp Glu Arg Asn His Arg Pro Gly Asp Glu Leu
900 905 910

Tyr Leu Pro Glu Leu Ala Ala Arg Trp Phe Glu Ala Asn Arg Pro Thr
 915 920 925
 Arg Pro Thr Leu Thr Ile Thr Glu Asp Val Ala Arg Thr Ala Asn Leu
 930 935 940
 Ala Ile Glu Leu Asp Ser Ala Thr Asp Val Gly Arg Ala Cys Ala Gly
 945 950 955 960
 Cys Arg Val Thr Pro Gly Val Val Gln Tyr Gln Phe Thr Ala Gly Val
 965 970 975
 Pro Gly Ser Gly Lys Ser Arg Ser Ile Thr Gln Ala Asp Val Asp Val
 980 985 990
 Val Val Val Pro Thr Arg Glu Leu Arg Asn Ala Trp Arg Arg Arg Gly
 995 1000 1005
 Phe Ala Ala Phe Thr Pro His Thr Ala Ala Arg Val Thr Gln Gly Arg
 1010 1015 1020
 Arg Val Val Ile Asp Glu Ala Pro Ser Leu Pro Pro His Leu Leu Leu
 1025 1030 1035 1040
 Leu His Met Gln Arg Ala Ala Thr Val His Leu Leu Gly Asp Pro Asn
 1045 1050 1055
 Gln Ile Pro Ala Ile Asp Phe Glu His Ala Gly Leu Val Pro Ala Ile
 1060 1065 1070
 Arg Pro Asp Leu Gly Pro Thr Ser Trp Trp His Val Thr His Arg Trp
 1075 1080 1085
 Pro Ala Asp Val Cys Glu Leu Ile Arg Gly Ala Tyr Pro Met Ile Gln
 1090 1095 1100
 Thr Thr Ser Arg Val Leu Arg Ser Leu Phe Trp Gly Glu Pro Ala Val
 1105 1110 1115 1120
 Gly Gln Lys Leu Val Phe Thr Gln Ala Ala Lys Pro Ala Asn Pro Gly
 1125 1130 1135
 Ser Val Thr Val His Glu Ala Gln Gly Ala Thr Tyr Thr Glu Thr Thr
 1140 1145 1150
 Ile Ile Ala Thr Ala Asp Ala Arg Gly Leu Ile Gln Ser Ser Arg Ala
 1155 1160 1165
 His Ala Ile Val Ala Leu Thr Arg His Thr Glu Lys Cys Val Ile Ile
 1170 1175 1180
 Asp Ala Pro Gly Leu Leu Arg Glu Val Gly Ile Ser Asp Ala Ile Val
 1185 1190 1195 1200
 Asn Asn Phe Phe Leu Ala Gly Gly Glu Ile Gly His Gln Arg Pro Ser
 1205 1210 1215
 Val Ile Pro Arg Gly Asn Pro Asp Ala Asn Val Asp Thr Leu Ala Ala
 1220 1225 1230
 Phe Pro Pro Ser Cys Gln Ile Ser Ala Phe His Gln Leu Ala Glu Glu

1235					1240					1245				
Leu Gly His Arg Pro Val	1250	Pro Val Ala Ala Val	1255	Leu Pro Pro Cys Pro	1260									
Glu Leu Glu Gln Gly	1265	Leu Leu Tyr Leu Pro	1270	Gln Glu Leu Thr Thr Cys	1275				1280					
Asp Ser Val Val Thr Phe	1285	Glu Leu Thr Asp Ile	1290	Val His Cys Arg Met	1295									
Ala Ala Pro Ser Gln Arg	1300	Lys Ala Val Leu Ser Thr	1305	Leu Val Gly Arg	1310									
Tyr Gly Gly Arg Thr Lys	1315	Leu Tyr Asn Ala Ser His	1320	Ser Asp Val Arg	1325									
Asp Ser Leu Ala Arg Phe	1330	Ile Pro Ala Ile Gly	1335	Pro Val Gln Val Thr	1340									
Thr Cys Glu Leu Tyr Glu	1345	Leu Val Glu Ala Met	1350	Val Glu Lys Gly Gln	1355				1360					
Asp Gly Ser Ala Val Leu	1365	Glu Leu Asp Leu Cys	1370	Asn Arg Asp Val Ser	1375									
Arg Ile Thr Phe Phe Gln	1380	Lys Asp Cys Asn Lys	1385	Phe Thr Thr Gly Glu	1390									
Thr Ile Ala His Gly Lys	1395	Val Gly Gln Gly Ile	1400	Ser Ala Trp Ser Lys	1405									
Thr Phe Cys Ala Leu Phe	1410	Gly Pro Trp Phe Arg	1415	Ala Ile Glu Lys Ala	1420									
Ile Leu Ala Leu Leu Pro	1425	Gln Gly Val Phe Tyr	1430	Gly Asp Ala Phe Asp	1435				1440					
Asp Thr Val Phe Ser Ala	1445	Ala Val Ala Ala Lys	1450	Ala Ser Met Val	1455									
Phe Glu Asn Asp Phe Ser	1460	Glu Phe Asp Ser Thr	1465	Gln Asn Asn Phe Ser	1470									
Leu Gly Leu Glu Cys Ala	1475	Ile Met Glu Glu Cys	1480	Gly Met Pro Gln Trp	1485									
Leu Ile Arg Leu Tyr His	1490	Leu Ile Arg Ser Ala	1495	Trp Ile Leu Gln Ala	1500									
Pro Lys Glu Ser Leu Arg	1505	Gly Phe Trp Lys Lys	1510	His Ser Gly Glu Pro	1515				1520					
Gly Thr Leu Leu Trp Asn	1525	Thr Val Trp Asn Met	1530	Ala Val Ile Thr His	1535									
Cys Tyr Asp Phe Arg Asp	1540	Phe Gln Val Ala Ala	1545	Phe Lys Gly Asp Asp	1550									
Ser Ile Val Leu Cys Ser	1555	Glu Tyr Arg Gln Ser	1560	Pro Gly Ala Ala Val	1565									

Leu Ile Ala Gly Cys Gly Leu Lys Leu Lys Val Asp Phe Arg Pro Ile
 1570 1575 1580
 Gly Leu Tyr Ala Gly Val Val Val Ala Pro Gly Leu Gly Ala Leu Pro
 1585 1590 1595 1600
 Asp Val Val Arg Phe Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro
 1605 1610 1615
 Gly Pro Glu Arg Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu
 1620 1625 1630
 Arg Lys Leu Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg
 1635 1640 1645
 Val Tyr Gly Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu
 1650 1655 1660
 Gln Ala Val Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro
 1665 1670 1675 1680
 Val Leu Asp Leu Thr Asn Ser Ile Leu Cys Arg Val Glu
 1685 1690

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 660 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Arg Pro Arg Pro Ile Leu Leu Leu Leu Leu Met Phe Leu Pro Met
 1 5 10 15
 Leu Pro Ala Pro Pro Pro Gly Gln Pro Ser Gly Arg Arg Arg Gly Arg
 20 25 30
 Arg Ser Gly Gly Ser Gly Gly Gly Phe Trp Gly Asp Arg Val Asp Ser
 35 40 45
 Gln Pro Phe Ala Ile Pro Tyr Ile His Pro Thr Asn Pro Phe Ala Pro
 50 55 60
 Asp Val Thr Ala Ala Ala Gly Ala Gly Pro Arg Val Arg Gln Pro Ala
 65 70 75 80
 Arg Pro Leu Gly Ser Ala Trp Arg Asp Gln Ala Gln Arg Pro Ala Val
 85 90 95
 Ala Ser Arg Arg Arg Pro Thr Thr Ala Gly Ala Ala Pro Leu Thr Ala
 100 105 110
 Val Ala Pro Ala His Asp Thr Pro Pro Val Pro Asp Val Asp Ser Arg
 115 120 125
 Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr
 130 135 140

Ser Ser Val Ala Thr Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu
145 150 155 160

Ser Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala
165 170 175

Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Ala Arg Ala Thr Ile
180 185 190

Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser
195 200 205

Ile Ser Phe Trp Pro Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met
210 215 220

Asn Ser Ile Thr Ser Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile
225 230 235 240

Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn Gln
245 250 255

Gly Trp Arg Ser Val Glu Thr Ser Gly Val Ala Glu Glu Glu Ala Thr
260 265 270

Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Leu Val Asn Ser Tyr
275 280 285

Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu Asp Phe Ala Leu
290 295 300

Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn Thr Asn Thr Arg Val
305 310 315 320

Ser Arg Tyr Ser Ser Thr Ala Arg His Arg Leu Arg Arg Gly Ala Asp
325 330 335

Gly Thr Ala Glu Leu Thr Thr Thr Ala Ala Thr Arg Phe Met Lys Asp
340 345 350

Leu Tyr Phe Thr Ser Thr Asn Gly Val Gly Glu Ile Gly Arg Gly Ile
355 360 365

Ala Leu Thr Leu Phe Asn Leu Ala Asp Thr Leu Leu Gly Gly Leu Pro
370 375 380

Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro
385 390 395 400

Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val
405 410 415

Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp
420 425 430

Leu Gly Glu Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln His Glu
435 440 445

Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val
450 455 460

Leu Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr
 465 470 475 480
 Asp Gln Ser Thr Tyr Gly Ser Ser Thr Gly Pro Val Tyr Val Ser Asp
 485 490 495
 Ser Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg
 500 505 510
 Ser Leu Asp Trp Thr Lys Val Thr Leu Asp Gly Arg Pro Leu Ser Thr
 515 520 525
 Ile Gln Gln Tyr Ser Lys Thr Phe Phe Val Leu Pro Leu Arg Gly Lys
 530 535 540
 Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn
 545 550 555 560
 Tyr Asn Thr Thr Ala Ser Asp Gln Leu Leu Val Glu Asn Ala Ala Gly
 565 570 575
 His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro
 580 585 590
 Val Ser Ile Ser Ala Val Ala Val Leu Ala Pro His Ser Ala Leu Ala
 595 600 605
 Leu Leu Glu Asp Thr Leu Asp Tyr Pro Ala Arg Ala His Thr Phe Asp
 610 615 620
 Asp Phe Cys Pro Glu Cys Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe
 625 630 635 640
 Gln Ser Thr Val Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys
 645 650 655
 Thr Arg Glu Leu
 660

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 123 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Asn Asn Met Ser Phe Ala Ala Pro Met Gly Ser Arg Pro Cys Ala
 1 5 10 15
 Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro
 20 25 30
 Arg His Arg Pro Val Ser Arg Leu Ala Ala Val Val Gly Gly Ala Ala
 35 40 45
 Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser Pro
 50 55 60

Ser Gln Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met Ser
65 70 75 80

Pro Leu Arg Pro Gly Leu Asp Leu Val Phe Ala Asn Pro Pro Asp His
85 90 95

Ser Ala Pro Leu Gly Val Thr Arg Pro Ser Ala Pro Pro Leu Pro His
100 105 110

Val Val Asp Leu Pro Gln Leu Gly Pro Arg Arg
115 120

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Composite Mexico strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCCATGGAGG CCCACCAAGT CATTAAAGGCT CCTGGCATCA CTACTGCTAT TGAGCAAGCA	60
GCTCTAGCAG CGGCCAACTC CGCCCTTGCG AATGCTGTGG TGGTCCGGCC TTTCCTTTCC	120
CATCAGCAGG TTGAGATCCT TATAAATCTC ATGCAACCTC GGCAGCTGGT GTTTCGTCCT	180
GAGGTTTTTT GGAATCACCC GATTCAACGT GTTATACATA ATGAGCTTGA GCAGTATTGC	240
CGTGCTCGCT CGGGTCGCTG CCTTGAGATT GGAGCCCACC CACGCTCCAT TAATGATAAT	300
CCTAATGTCC TCCATCGCTG CTTTCTCCAC CCCGTCGGCC GGGATGTTCA GCGCTGGTAC	360
ACAGCCCCGA CTAGGGGACC TGCGGCGAAC TGTCGCCGCT CGGCACTTCG TGGTCTGCCA	420
CCAGCCGACC GCACTTACTG TTTTGATGGC TTTGCCGGCT GCCGTTTTGC CGCCGAGACT	480
GGTGTGGCTC TCTATTCTCT CCATGACTTG CAGCCGGCTG ATGTTGCCGA GGCGATGGCT	540
CGCCACGGCA TGACCCGCCT TTATGCAGCT TTCCACTTGC CTCCAGAGGT GTCCTGCCT	600
CCTGGCACCT ACCGGACATC ATCCTACTTG CTGATCCACG ATGGTAAGCG CGCGGTTGTC	660
ACTTATGAGG GTGACACTAG CGCCGGTTAC AATCATGATG TTGCCACCCT CCGCACATGG	720
ATCAGGACAA CTAAGGTTGT GGGTGAACAC CCTTTGGTGA TCGAGCGGGT GCGGGGTATT	780
GGCTGTCACT TTGTGTTGTT GATCACTGCG GCCCCTGAGC CCTCCCCGAT GCCCTACGTT	840
CCTTACCCGC GTTCGACGGA GGTCTATGTC CGGTCTATCT TTGGGCCCCG CGGGTCCCCG	900

TCGCTGTTCC	CGACCGCTTG	TGCTGTCAAG	TCCACTTTTC	ACGCCGTCCC	CACGCACATC	960
TGGGACCGTC	TCATGCTCTT	TGGGGCCACC	CTCGACGACC	AGGCCTTTTG	CTGCTCCAGG	1020
CTTATGACGT	ACCTTCGTGG	CATTAGCTAT	AAGGTAAGTG	TGGGTGCCCT	GGTCGCTAAT	1080
GAAGGCTGGA	ATGCCACCGA	GGATGCGCTC	ACTGCAGTTA	TTACGGCGGC	TTACCTCACA	1140
ATATGTCATC	AGCGTTATTT	GCGGACCCAG	GCGATTCTTA	AGGGCATGCG	CCGGCTTGAG	1200
CTTGAACATG	CTCAGAAATT	TATTTACAGC	CTCTACAGCT	GGCTATTTGA	GAAGTCAGGT	1260
CGTGATTACA	TCCCAGGCCG	CCAGCTGCAG	TTCTACGCTC	AGTGCCGCCG	CTGGTTATCT	1320
GCCGGGTTCC	ATCTCGACCC	CCGCACCTTA	GTTTTTGATG	AGTCAGTGCC	TTGTAGCTGC	1380
CGAACCACCA	TCCGGCGGAT	CGCTGGAAAA	TTTGCTGTT	TTATGAAGTG	GCTCGGTCAG	1440
GAGTGTTCCT	GTTTCCTCCA	GCCCGCCGAG	GGGCTGGCGG	GCGACCAAGG	TCATGACAAT	1500
GAGGCCTATG	AAGGCTCTGA	TGTTGATACT	GCTGAGCCTG	CCACCCTAGA	CATTACAGGC	1560
TCATACATCG	TGGATGGTCG	GTCTCTGCAA	ACTGTCTATC	AAGCTCTCGA	CCTGCCAGCT	1620
GACCTGGTAG	CTCGCGCAGC	CCGACTGTCT	GCTACAGTTA	CTGTTACTGA	AACCTCTGGC	1680
CGTCTGGATT	GCCAAACAAT	GATCGGCAAT	AAGACTTTTC	TCACTACCTT	TGTTGATGGG	1740
GCACGCCTTG	AGGTTAACGG	GCCTGAGCAG	CTTAACCTCT	CTTTTGACAG	CCAGCAGTGT	1800
AGTATGGCAG	CCGGCCCGTT	TTGCCTCACC	TATGCTGCCG	TAGATGGCGG	GCTGGAAGTT	1860
CATTTTTCCA	CCGCTGGCCT	CGAGAGCCGT	GTTGTTTTCC	CCCCTGGTAA	TGCCCCGACT	1920
GCCCCGCCGA	GTGAGGTCAC	CGCCTTCTGC	TCAGCTCTTT	ATAGGCACAA	CCGGCAGAGC	1980
CAGCGCCAGT	CGTTATTGG	TAGTTTGTGG	CTGCACCCTG	AAGGTTTGCT	CGGCCTGTTC	2040
CCGCCCTTTT	CACCCGGGCA	TGAGTGGCGG	TCTGCTAACC	CATTTTGCGG	CGAGAGCACG	2100
CTCTACACCC	GCACTTGGTC	CACAATTACA	GACACACCTT	TAAGTGTGCG	GCTAATTTCC	2160
GGTCATTTGG	ATGCTGCTCC	CCACTCGGGG	GGGCCACCTG	CTACTGCCAC	AGGCCCTGCT	2220
GTAGGCTCGT	CTGACTCTCC	AGACCCTGAC	CCGCTACCTG	ATGTTACAGA	TGGCTCACGC	2280
CCCTCTGGGG	CCCGTCCGGC	TGGCCCCAAC	CCGAATGGCG	TTCCGCAGCG	CCGCTTACTA	2340
CACACCTACC	CTGACGGCGC	TAAGATCTAT	GTCGGCTCCA	TTTTCGAGTC	TGAGTGCACC	2400
TGGCTTGTC	ACGCATCTAA	CGCCGGCCAC	CGCCCTGGTG	GCGGGCTTTG	TCATGCTTTT	2460
TTTCAGCGTT	ACCTGATTC	GTTTGACGCC	ACCAAGTTTG	TGATGCGTGA	TGGTCTTGCC	2520
GCGTATACCC	TTACACCCCG	GCCGATCATT	CATGCGGTGG	CCCCGGACTA	TCGATTGGAA	2580
CATAACCCCA	AGAGGCTCGA	GGCTGCCTAC	CGCGAGACTT	GCGCCCGCCG	AGGCACTGCT	2640
GCCTATCCAC	TCTTAGGCGC	TGGCATTAC	CAGGTGCCTG	TTAGTTTGAG	TTTTGATGCC	2700
TGGGAGCGGA	ACCACCGCCC	GTTTGACGAG	CTTACCTAA	CAGAGCTGGC	GGCTCGGTGG	2760

TTTGAATCCA	ACCGCCCCGG	TCAGCCCACG	TTGAACATAA	CTGAGGATAC	CGCCCGTGCG	2820
GCCAACCTGG	CCCTGGAGCT	TGACTCCGGG	AGTGAAGTAG	GCCGCGCATG	TGCCGGGTGT	2880
AAAGTCGAGC	CTGGCGTTGT	GCGGTATCAG	TTTACAGCCG	GTGTCCCCGG	CTCTGGCAAG	2940
TCAAAGTCCG	TGCAACAGGC	GGATGTGGAT	GTTGTTGTTG	TGCCCCACTCG	CGAGCTTCGG	3000
AACGCTTGGC	GGCGCCGGGG	CTTTGCGGCA	TTCACTCCGC	ACACTGCGGC	CCGTGTCACT	3060
AGCGGCCGTA	GGGTTGTCAT	TGATGAGGCC	CCTTCGCTCC	CCCCACACTT	GCTGCTTTTA	3120
CATATGCAGC	GTGCTGCATC	TGTGCACCTC	CTTGGGGACC	CGAATCAGAT	CCCCGCCATA	3180
GATTTTGAGC	ACACCGGTCT	GATTCCAGCA	ATACGGCCGG	AGTTGGTCCC	GACTTCATGG	3240
TGGCATGTCA	CCCACCGTTG	CCCTGCAGAT	GTCTGTGAGT	TAGTCCGTGG	TGCTTACCCT	3300
AAAATCCAGA	CTACAAGTAA	GGTGCTCCGT	TCCCTTTTCT	GGGGAGAGCC	AGCTGTCCGG	3360
CAGAAGCTAG	TGTTACACA	GGCTGCTAAG	GCCGCGCACC	CCGGATCTAT	AACGGTCCAT	3420
GAGGCCCAGG	GTGCCACTTT	TACCACTACA	ACTATAATTG	CAACTGCAGA	TGCCCGTGGC	3480
CTCATACAGT	CCTCCCGGGC	TCACGCTATA	GTTGCTCTCA	CTAGGCATAC	TGAAAAATGT	3540
GTTATACTTG	ACTCTCCCGG	CCTGTTGCGT	GAGGTGGGTA	TCTCAGATGC	CATTGTTAAT	3600
AATTTCTTCC	TTTCGGGTGG	CGAGGTTGGT	CACCAGAGAC	CATCGGTCAT	TCCGCGAGGC	3660
AACCCTGACC	GCAATGTTGA	CGTGCTTGCG	GCGTTTCCAC	CTTCATGCCA	AATAAGCGCC	3720
TTCCATCAGC	TTGCTGAGGA	GCTGGGCCAC	CGGCCGGCGC	CGGTGGCGGC	TGTGCTACCT	3780
CCCTGCCCTG	AGCTTGAGCA	GGCCTTCTC	TATCTGCCAC	AGGAGCTAGC	CTCCTGTGAC	3840
AGTGTTGTGA	CATTTGAGCT	AACTGACATT	GTGCACTGCC	GCATGGCGGC	CCCTAGCCAA	3900
AGGAAAGCTG	TTTTGTCCAC	GCTGGTAGGC	CGGTATGGCA	GACGCACAAG	GCTTTATGAT	3960
GCGGGTCACA	CCGATGTCCG	CGCCTCCCTT	GCGCGCTTTA	TTCCCACTCT	CGGGCGGGTT	4020
ACTGCCACCA	CCTGTGAACT	CTTTGAGCTT	GTAGAGGCGA	TGGTGGAGAA	GGGCCAAGAC	4080
GGTTCAGCCG	TCCTCGAGTT	GGATTTGTGC	AGCCGAGATG	TCTCCCGCAT	AACCTTTTTT	4140
CAGAAGGATT	GTAACAAGTT	CACGACCGGC	GAGACAATTG	CGCATGGCAA	AGTCGGTCAG	4200
GGTATCTTCC	GCTGGAGTAA	GACGTTTTGT	GCCCTGTTTG	GCCCCTGGTT	CCGTGCGATT	4260
GAGAAGGCTA	TTCTATCCCT	TTTACCACAA	GCTGTGTTCT	ACGGGGATGC	TTATGACGAC	4320
TCAGTATTCT	CTGCTGCCGT	GGCTGGCGCC	AGCCATGCCA	TGGTGTTTGA	AAATGATTTT	4380
TCTGAGTTTG	ACTCGACTCA	GAATAACTTT	TCCCTAGGTC	TTGAGTGCGC	CATTATGGAA	4440
GAGTGTGGTA	TGCCCCAGTG	GCTTGTCAGG	TTGTACCATG	CCGTCCGGTC	GGCGTGGATC	4500
CTGCAGGCCC	CAAAGAGTC	TTTGAGAGGG	TTCTGGAAGA	AGCATTCTGG	TGAGCCGGGC	4560
AGCTTGCTCT	GGAATACGGT	GTGGAACATG	GCAATCATTG	CCCATTGCTA	TGAGTTCCGG	4620

GACCTCCAGG	TTGCCGCCTT	CAAGGGCGAC	GACTCGGTCTG	TCCTCTGTAG	TGAATACCGC	4680
CAGAGCCCAG	GCGCCGGTTC	GCTTATAGCA	GGCTGTGGTT	TGAAGTTGAA	GGCTGACTTC	4740
CGGCCGATTG	GGCTGTATGC	CGGGGTGTGC	GTCGCCCCGG	GGCTCGGGGC	CCTACCCGAT	4800
GTCGTTTCGAT	TCGCCGGACG	GCTTTCGGAG	AAGAACTGGG	GGCCTGATCC	GGAGCGGGCA	4860
GAGCAGCTCC	GCCTCGCCGT	GCAGGATTTT	CTCCGTAGGT	TAACGAATGT	GGCCCAGATT	4920
TGTGTTGAGG	TGGTGTCTAG	AGTTTACGGG	GTTTCCCCGG	GTCTGGTTCA	TAACCTGATA	4980
GGCATGCTCC	AGACTATTGG	TGATGGTAAG	GCGCATTTTA	CAGAGTCTGT	TAAGCCTATA	5040
CTTGACCTTA	CACACTCAAT	TATGCACCGG	TCTGAATGAA	TAACATGTGG	TTTGCTGCGC	5100
CCATGGGTTT	GCCACCATGC	GCCCTAGGCC	TCTTTTGCTG	TTGTTCTCTT	TGTTTCTGCC	5160
TATGTTGCCC	GCGCCACCGA	CCGGTCAGCC	GTCTGGCCGC	CGTCGTGGGC	GGCGCAGCGG	5220
CGGTACCGGC	GGTGGTTTCT	GGGGTGACCG	GGTTGATTCT	CAGCCCTTCG	CAATCCCCTA	5280
TATTCATCCA	ACCAACCCCT	TTGCCCCAGA	CGTTGCCGCT	GCGTCCGGGT	CTGGACCTCG	5340
CCTTCGCCAA	CCAGCCCGGC	CACTTGGCTC	CACTTGGCGA	GATCAGGCCC	AGCGCCCCCT	5400
CGCTGCCTCC	CGTCGCCGAC	CTGCCACAGC	CGGGGCTGCG	GCGCTGACGG	CTGTGGCGCC	5460
TGCCCATGAC	ACCTCACCCG	TCCCGGACGT	TGATTCTCGC	GGTGCAATTC	TACGCCGCCA	5520
GTATAATTTG	TCTACTTCAC	CCCTGACATC	CTCTGTGGCC	TCTGGCACTA	ATTTAGTCCT	5580
GTATGCAGCC	CCCCTTAATC	CGCCTCTGCC	GCTGCAGGAC	GGTACTAATA	CTCACATTAT	5640
GGCCACAGAG	GCCTCCAATT	ATGCACAGTA	CCGGGTGTCG	CGCGCTACTA	TCCGTTACCG	5700
GCCCCTAGTG	CCTAATGCAG	TTGGAGGCTA	TGCTATATCC	ATTTCTTTCT	GGCCTCAAAC	5760
AACCACAACC	CCTACATCTG	TTGACATGAA	TTCCATTACT	TCCACTGATG	TCAGGATTCT	5820
TGTTCAACCT	GGCATAGCAT	CTGAATTGGT	CATCCCAAGC	GAGCGCCTTC	ACTACGCCAA	5880
TCAAGGTTGG	CGCTCGGTTG	AGACATCTGG	TGTTGCTGAG	GAGGAAGCCA	CCTCCGGTCT	5940
TGTCATGTTA	TGCATACATG	GCTCTCCAGT	TAACTCCTAT	ACCAATACCC	CTTATACCGG	6000
TGCCCTTGGC	TACTGGAAT	TTGCCTTAGA	GCTTGAGTTT	CGCAATCTCA	CCACCTGTAA	6060
CACCAATACA	CGTGTGTCCC	GTTACTCCAG	CACTGCTCGT	CACTCCGCCC	GAGGGGCCGA	6120
CGGGACTGCG	GAGCTGACCA	CAACTGCAGC	CACCAGGTTT	ATGAAAGATC	TCCACTTTAC	6180
CGGCCTTAAT	GGGGTAGGTG	AAGTCGGCCG	CGGGATAGCT	CTAACATTAC	TTAACCTTGC	6240
TGACACGCTC	CTCGCGGGGC	TCCCGACAGA	ATTAATTTTC	TCGGCTGGCG	GGCAACTGTT	6300
TTATTCCCGC	CCGGTTGTCT	CAGCCAATGG	CGAGCCAACC	GTGAAGCTCT	ATACATCAGT	6360
GGAGAATGCT	CAGCAGGATA	AGGGTGTTGC	TATCCCCCAC	GATATCGATC	TTGGTGATTCT	6420
GCGTGTGGTC	ATTCAGGATT	ATGACAACCA	GCATGAGCAG	GATCGGCCCA	CCCCGTGCGC	6480

TGCGCCATCT CGGCCTTTTT CTGTTCTCCG AGCAAATGAT GTACTTTGGC TGTCCCTCAC 6540
TGCAGCCGAG TATGACCAGT CCACTTACGG GTCGTCAACT GGCCCGGTTT ATATCTCGGA 6600
CAGCGTGACT TTGGTGAATG TTGCGACTGG CGCGCAGGCC GTAGCCCGAT CGCTTGACTG 6660
GTCCAAAGTC ACCCTCGACG GCGGGCCCT CCCGACTGTT GAGCAATATT CCAAGACATT 6720
CTTTGTGCTC CCCCTTCGTG GCAAGCTCTC CTTTGGGAG GCCGGCACAA CAAAAGCAGG 6780
TTATCCTTAT AATTATAATA CTACTGCTAG TGACCAGATT CTGATTGAAA ATGCTGCCGG 6840
CCATCGGGTC GCCATTTCAA CCTATACCAC CAGGCTTGGG GCCGGTCCGG TCGCCATTTTC 6900
TGCGGCCGCG GTTTTGGCTC CACGCTCCGC CCTGGCTCTG CTGGAGGATA CTTTGTATTA 6960
TCCGGGGCGG GCGCACACAT TTGATGACTT CTGCCCTGAA TGCCGCGCTT TAGGCCTCCA 7020
GGGTTGTGCT TTCCAGTCAA CTGTCGCTGA GCTCCAGCGC CTAAAGTTA AGGTGGGTAA 7080
AACTCGGGAG TTGTAGTTTA TTTGGCTGTG CCCACCTACT TATATCTGCT GATTTCCTTT 7140
ATTCCTTTT TCTCGGTCCC GCGCTCCCTG A 7171

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1575 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: T: Mexican strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTTGCGTGAG GTGGGTATCT CAGATGCCAT TGTTAATAAT TTCTTCCTTT CGGGTGGCGA 60
GGTTGGTCAC CAGAGACCAT CGGTCATTCC GCGAGGCAAC CCTGACCGCA ATGTTGACGT 120
GCTTGCGGCG TTTCCACCTT CATGCCAAAT AAGCGCCTTC CATCAGCTTG CTGAGGAGCT 180
GGGCCACCGG CCGGCGCCGG TGGCGGCTGT GCTACCTCCC TGCCCTGAGC TTGAGCAGGG 240
CCTTCTCTAT CTGCCACAGG AGCTAGCCTC CTGTGACAGT GTTGTGACAT TTGAGCTAAC 300
TGACATTGTG CACTGCCGCA TGGCGGCCCC TAGCCAAAGG AAAGCTGTTT TGTCCACGCT 360
GGTAGGCCGG TATGGCAGAC GCACAAGGCT TTATGATGCG GGTCACACCG ATGTCCGCGC 420
CTCCCTTGCG CGCTTTATTC CCACTCTCGG GCGGGTTACT GCCACCACCT GTGAACTCTT 480
TGAGCTTGTA GAGGCGATGG TGGAGAAGGG CCAAGACGGT TCAGCCGTCC TCGAGTTGGA 540

TTTGTGCAGC CGAGATGTCT CCCGCATAAC CTTTTTCCAG AAGGATTGTA ACAAGTTCAC	600
GACCGGCGAG ACAATTGCGC ATGGCAAAGT CGGTCAGGGT ATCTTCCGCT GGAGTAAGAC	660
CTTTTGTGCC CTGTTTGGCC CCTGGTTCGG TCGATTGAG AAGGCTATTC TATCCCTTTT	720
ACCACAAGCT GTGTTCTACG GGGATGCTTA TGACGACTCA GTATTCTCTG CTGCCGTGGC	780
TGGCGCCAGC CATGCCATGG TGTTTGAAAA TGATTTTCT GAGTTTGA CTGACTCAGAA	840
TAACTTTTCC CTAGGTCTTG AGTGCGCCAT TATGGAAGAG TGTGGTATGC CCCAGTGGCT	900
TGTCAGGTTG TACCATGCCG TCCGGTCGGC GTGGATCCTG CAGGCCCCAA AAGAGTCTTT	960
GAGAGGGTTC TGGAAGAAGC ATTCTGGTGA GCCGGGCACG TTGCTCTGGA ATACGGTGTG	1020
GAACATGGCA ATCATTGCCC ATTGCTATGA GTTCCGGGAC CTCCAGGTTG CCGCCTTCAA	1080
GGGCGACGAC TCGGTCGTCC TCTGTAGTGA ATACCGCCAG AGCCCAGGCG CCGGTTGCT	1140
TATAGCAGGC TGTGGTTTGA AGTTGAAGGC TGA CTCCGG CCGATTGGGC TGTATGCCGG	1200
GGTTGTCGTC GCCCCGGGGC TCGGGGCCCT ACCCGATGTC GTTCGATTG CCGGACGGCT	1260
TTCGGAGAAG AACTGGGGGC CTGATCCGGA GCGGGCAGAG CAGCTCCGCC TCGCCGTGCA	1320
GGATTTCTC CGTAGGTAA CGAATGTGGC CCAGATTTGT GTTGAGGTGG TGTCTAGAGT	1380
TTACGGGGTT TCCCCGGGTC TGGTTCATAA CCTGATAGGC ATGCTCCAGA CTATTGGTGA	1440
TGGTAAGGCG CATTTTACAG AGTCTGTTAA GCCTATACTT GACCTTACAC ACTCAATTAT	1500
GCACCGGTCT GAATGAATAA CATGTGGTTT GCTGCGCCCA TGGGTTCGCC ACCATGCGCC	1560
CTAGGCCTCT TTTGC	1575

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Tashkent strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGGGCCCCGT ACAGGTCACA ACCTGTGAGT TGTACGAGCT AGTGGAGGCC ATGGTCGAGA	60
AAGGCCAGGA TGGCTCCGCC GTCCTTGAGC TCGATCTCTG CAACCGTGAC GTGTCCAGGA	120
TCACCTTTTT CCAGAAAGAT TGCAATAAGT TCACCACGGG AGAGACCATC GCCCATGGTA	180

```

AAGTGGGCCA GGGCATTTCG GCCTGGAGTA AGACCTTCTG TGCCCTTTTC GGCCCTGGT      240
TCCGTGCTAT TGAGAAGGCT ATTCTGGCCC TGCTCCCTCA GGGTGTGTTT TATGGGGATG      300
CCTTTGATGA CACCGTCTTC TCGGCGCGTG TGGCCGCAGC AAAGGCGTCC ATGGTGTTTG      360
AGAATGACTT TTCTGAGTTT GACTCCACCC AGAATAATTT TTCCCTGGGC CTAGAGTGTG      420
CTATTATGGA GAAGTGTGGG ATGCCGAAGT GGCTCATCCG CTTGTACCAC CTTATAAGGT      480
CTGCGTGGAT CCTGCAGGCC CCGAAGGAGT CCCTGCGAGG GTGTTGGAAG AAACACTCCG      540
GTGAGCCCGG CACTCTTCTA TGGAATACTG TCTGGAACAT GGCCGTTATC ACCCATTGTT      600
ACGATTTCCG CGATTTGCAG GTGGCTGCCT TTAAAGGTGA TGATTCGATA GTGCTTTGCA      660
GTGAGTACCG TCAGAGTCCA GGGGCTGCTG TCCTGATTGC TGGCTGTGGC TTAAAGCTGA      720
AGGTGGGTTT CCGTCCGATT GGTTTGTATG CAGGTGTTGT GGTGACCCCC GGCCTTGGCG      780
CGCTTCCCGA CGTCGTGCGC TTGTCCGGCC GGCTTACTGA GAAGAATTGG GGCCCTGGCC      840
CTGAGCGGGC GGAGCAGCTC CGCCTTGCTG TGCG                                  874

```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 449 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone 406.4-2 cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..100

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

C GCC AAC CAG CCC GGC CAC TTG GCT CCA CTT GGC GAG ATC AGG CCC      46
Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro
  1             5             10             15

AGC GCC CCT CCG CTG CCT CCC GTC GCC GAC CTG CCA CAG CCG GGG CTG      94
Ser Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu
                20             25             30

CGG CGC TGACGGCTGT GGCGCCTGCC CATGACACCT CACCCGTCCC GGACGTTGAT      150
Arg Arg

TCTCGCGGTG CAATTCTACG CCGCCAGTAT AATTTGTCTA CTTACCCCT GACATCCTCT      210

```

GTGGCCTCTG GCACTAATTT AGTCCTGTAT GCAGCCCCCC TTAATCCGCC TCTGCCGCTG 270
 CAGGACGGTA CTAATACTCA CATTATGGCC ACAGAGGCCT CCAATTATGC ACAGTACCGG 330
 GTTGCCCGCG CTACTATCCG TTACCGGCCC CTAGTGCCTA ATGCAGTTGG AGGCTATGCT 390
 ATATCCATTT CTTTCTGGCC TCAAACAACC ACAACCCCTA CATCTGTTGA CATGAATTC 449

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser
 1 5 10 15

Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg
 20 25 30

Arg

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 130 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: Clone 406.3-2

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 5..130

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAT ACT TTT GAT TAT CCG GGG CGG GCG CAC ACA TTT GAT GAC TTC TGC 49
 Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys
 1 5 10 15

CCT GAA TGC CGC GCT TTA GGC CTC CAG GGT TGT GCT TTC CAG TCA ACT 97
 Pro Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr
 20 25 30

GTC GCT GAG CTC CAG CGC CTT AAA GTT AAG GTT 130

Val Ala Glu Leu Gln Arg Leu Lys Val Lys Val
 35 40

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys Pro
 1 5 10 15
 Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val
 20 25 30
 Ala Glu Leu Gln Arg Leu Lys Val Lys Val
 35 40

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: 406.4-2 epitope - Mexican strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser
 1 5 10 15
 Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg
 20 25 30
 Arg

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: 406.4-2 epitope - Burma strain
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala	Asn	Pro	Pro	Asp	His	Ser	Ala	Pro	Leu	Gly	Val	Thr	Arg	Pro	Ser
1				5					10					15	
Ala	Pro	Pro	Leu	Pro	His	Val	Val	Asp	Leu	Pro	Gln	Leu	Gly	Pro	Arg
			20					25					30		
Arg															

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown
 (ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (vi) ORIGINAL SOURCE:
 (C) INDIVIDUAL ISOLATE: 406.3-2 epitope - Mexican strain
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Thr	Phe	Asp	Tyr	Pro	Gly	Arg	Ala	His	Thr	Phe	Asp	Asp	Phe	Cys	Pro
1				5					10					15	
Glu	Cys	Arg	Ala	Leu	Gly	Leu	Gln	Gly	Cys	Ala	Phe	Gln	Ser	Thr	Val
			20					25					30		
Ala	Glu	Leu	Gln	Arg	Leu	Lys	Val	Lys	Val						
			35				40								

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown
 (ii) MOLECULE TYPE: peptide
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 406.3-2 epitope - Burma strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Thr Leu Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro
1 5 10 15

Glu Cys Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val
20 25 30

Ala Glu Leu Gln Arg Leu Lys Met Lys Val
35 40